

## Picking SNPs Application to Association Studies

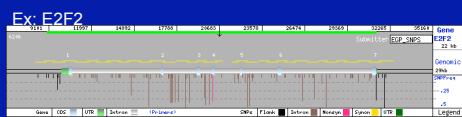
Dana Crawford, PhD

NIEHS  
Environmental Genome Project  
January 30, 2006

### Outline of Tutorial

- Concepts of tagSNPs
- LD and haplotype definitions
- Haplotype blocks and definitions
- Tools to identify tagSNPs

### Why Do We Need tagSNPs?



### Too Many SNPs to Genotype!

#### Whole Genome:

- 15,000,000 SNPs
- 6,000,000 SNPs  $\geq$  5% MAF

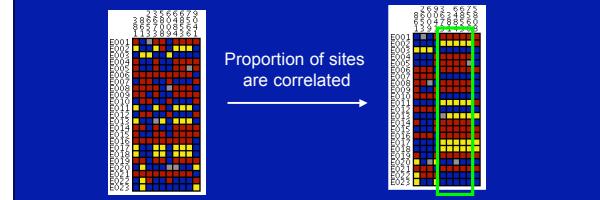
#### Average Gene:

- 26.5 kb
- 130 SNPs
- 44 SNPs  $\geq$  5% MAF

### SNPs Are Correlated (aka linkage disequilibrium)

"the nonindependence of alleles at different sites." Pritchard and Przeworski 2001

Genotype at one site can predict genotype at another site



### Measuring Pair-wise SNP Correlations

- SNP correlation described by linkage disequilibrium (LD)
- Pair-wise measures of LD: D' and r<sup>2</sup>

$$D = p_{AB} - p_A p_B; D' = D/D_{max}$$

Recombination

$$r^2 = \frac{D^2}{f(A_1)f(A_2)f(B_1)f(B_2)}$$

Power

### LD Statistics: Practical Uses

- r<sup>2</sup> is inversely related to power

$$\frac{1}{r^2}$$

1,000 cases	$\}$	$r^2=1.0$	1,250 cases	$\}$	$r^2 = 0.80$
1,000 controls			1,250 controls		

- D' is related to recombination history

$D' = 1$	no recombination
$D' < 1$	historical recombination

## Where to Find Population LD Statistics

For your gene or region of interest, search

- HapMap [www.hapmap.org](http://www.hapmap.org)
  - Perlegen [genome.perlegen.com](http://genome.perlegen.com)
  - SeattleSNPs PGA [pga.gs.washington.edu](http://pga.gs.washington.edu)
  - NIEHS SNPs [egp.gs.washington.edu](http://egp.gs.washington.edu)

## Where to Find Population LD Statistics

For your gene or region of interest, search

- HapMap [www.hapmap.org](http://www.hapmap.org)
  - Perlegen [genome.perlegen.com](http://genome.perlegen.com)
  - SeattleSNPs PGA [pga.gs.washington.edu](http://pga.gs.washington.edu)
  - NIEHS SNPs [egp.gs.washington.edu](http://egp.gs.washington.edu)

## Visualizing Pair-wise LD



**National Institute of Environmental Health Sciences**  
**Environmental Genome Project**  
**NIEHS SNPs**

**Search Site**

**Home**  
[Gene Nomenclature Form](#)

**Gene Targets**  
[Finished Genes Table](#) [Candidate Genes Table](#) [SNP Targets Table](#)

**Genotyping Resources**  
[Genotyping Backbone](#) [Genotyping Services](#) [Genotyping Data](#) [Genotype Download](#) [Submit New Data](#)

**Pathway Enrichment**  
[Basic Enrichment](#) [Cell Cycle Enrichment](#) [Dose Strand Break Repair Enrichment](#) [Transcription Coupled Repair Enrichment](#)

**Personnel**  
[Summary Information for Contributors](#) [Summary Data](#) [Data Access Requests](#)

**Usage Policy**  
[Usage Categories](#) [Usage Guidelines](#)

**Information**

Welcome to the NIEHS SNPs Program

## Introduction

The NIEHS Environmental Genome Project is a multidisciplinary endeavor, dedicated to improving our understanding of the relationship between environmental exposures, individual susceptibility variants in human genes, and disease risk in U.S. populations. The NIEHS SNPs Program at the University of Washington is targeted on the identification of single nucleotide polymorphisms (SNPs) in human genes that are associated with environmental exposures. The first phase of the effort is focused on finding common response variation (SNPs) in the human genome that are associated with environmental exposures. As these SNPs are identified, the project will map their locations to human genes that are apolaged in regulating human disease risk with environmental exposures.

## GeneSNPs Database

NEHS SNPs are available in the [GeneSNPs](#) database as well as the national database resource, [dbSNP](#). GeneSNPs provides a gene-centered map of the genome structure, coding sequences, and identified SNPs. dbSNP is a public SNP database maintained by the National Center for Biotechnology Information (NCBI). It contains a large collection of SNPs from many sources. GeneSNPs provides a graphical view of all available SNP data including allele frequencies and genotypes in select populations. The dbSNP page provides a detailed description of the polygenic sites needed to examine disease risk in human population studies.

## Polymorphism Analysis

Automated DNA sequencing is being used to identify and genotype SNPs in human candidate genes (see [PhyloPhen](#)). Candidate genes are being sequenced to identify common sequence variations for functional analysis. The first phase of the project will focus on identifying SNPs in human genes that are apolaged in individuals representative of the U.S. population (see [Sample Population Description Part 1](#)). Candidate genes are selected based on their known function and association with environmental exposures (see [Population Description Part 2](#)). All SNPs have been identified using high quality sequence data (G > T, A > C).

## Visualizing Pair-wise LD

National Institute of Environmental Health Sciences  
Environmental Genome Project

**NIEHS SNPs**

Search Site  Go

**Home**

**Gene Information Form**

**Gene Targets**

- > [Gene Targets Checklist](#)
- > [Printed Genes Table](#)
- > [Gene Targets Table](#)
- > [Proposed Genes Table](#)

**Genotyping Resources**

- > [General Information](#)
- > [General Methodology](#)
- > [Sample Preparation](#)
- > [Filtering & Mapping](#)
- > [Assessments](#)
- > [Data Sources](#)
- > [Cell Culture](#)
- > [Designing Assays](#)
- > [Experimental Design](#)
- > [Instrumentation](#)
- > [Methodology](#)
- > [Statistical Methods](#)
- > [Transcription Coupled Repair](#)

**Local Pedigree Input File**

ESG Printed Steve Pedigree Input File

ADH1C

None, use local site

AB0:B1

Raw Allele Percentage Integer

Cluster under Draw Trees For

NO THANKS

LD

P2

NO THANKS

LD

P2

Linkage Disequilibrium P-value

Rainbow

Black & White

Greyscale

WADH1D

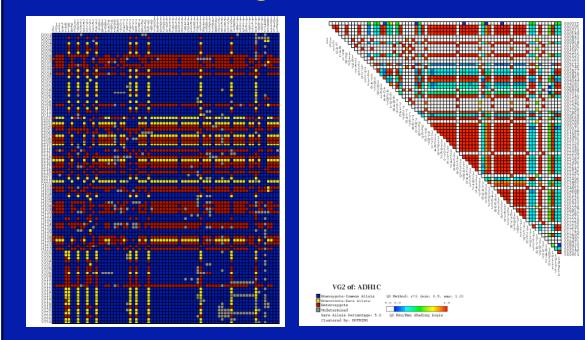
D4H

**Display Genotype Data: Visual Genotypes**

D4H

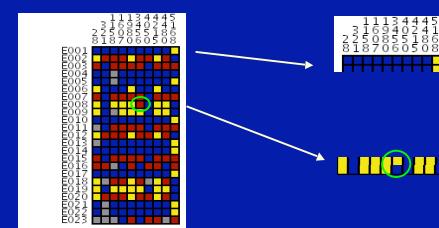
The display and interpretation of large genotype data sets can be simplified by using a visual genotype (VG) display. We have found it useful to present complete raw datasets of individuals' genotype data using a display format called a visual genotype (VG) (see Nickerson et al., *Nature Genetics*, 19(23-24), 1998, and Rieder et al., *Nature Genetics*, 22(9-10), 1999). This format presents all data in an array of samples (rows) x polymorphic sites (columns) and encodes each di-allelic polymorphism according to a general color scheme.

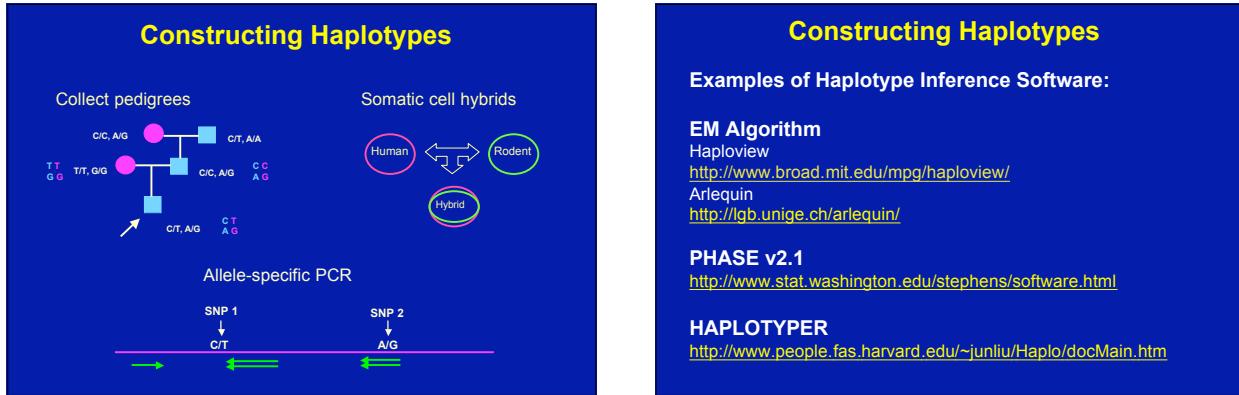
## Visualizing Pair-wise LD



## Multi-SNP Correlations (aka Haplotypes)

“...a unique combination of genetic markers present in a chromosome.” pg 57 in Hartl & Clark, 1997





## Haplotypes in NIEHS SNPs

- >550 genes re-sequenced  
Cell cycle, DNA repair/replication, apoptosis
- 2 DNA panels
  - Polymorphism Discovery Resource (PDR90)
  - Europeans, Africans, Hispanics, and Asians
- PHASEv2.0 results posted on website
- Interactive tool (VH1) to visualize and sort haplotypes

<http://egp.gs.washington.edu>

## Haplotypes in NIEHS SNPs

National Institute of Environmental Health Sciences Environmental Genome Project NIEHS SNPs

Welcome to the NIEHS SNPs Program

Introduction

The NIEHS Environmental Genome Project is a multi-disciplinary, collaborative effort focused on examining the relationships between environmental exposures, inter-individual sequence variation in human genes and disease susceptibility. The project will use high-throughput sequencing technologies to identify and characterize the systematic identification and genotyping of single nucleotide polymorphisms (SNPs) in environmental response genes. The ultimate goal of the project is to identify the genetic variants in human genes that are involved in environmental responses and disease susceptibility. In addition, the project will provide dense genetic maps of human genes that can be used to identify human disease risk with environmental exposures.

GeneSNPs Database

Latest Update to Finished Genes

CAB	December 20, 2005
MARCH2	December 19, 2005
ADH1A2	December 12, 2005
NEIL3	December 1, 2005
CHMP2B	November 30, 2005
CHMP2C	November 30, 2005
GPR93	November 30, 2005
RETIN	November 30, 2005
TSPY9K	November 22, 2005
ACVR1	November 22, 2005
MSX1	November 22, 2005
MSX2	November 22, 2005
SPRY1	November 4, 2005

Search Site Go

## Haplotypes in NIEHS SNPs

National Institute of Environmental Health Sciences Environmental Genome Project NIEHS SNPs

Finished Genes (alphabetical directory by Hugo Gene Name)

A| B| C| D| E| F| G| H| I| K| L| M| N| O| P| Q| R| S| T| U| Z

We welcome input in setting priorities for requesting target genes from NIEHS-funded investigators. To suggest a priority, please submit your request to Debbie Nickerson ([debnick@washington.edu](mailto:debnick@washington.edu)) OR postal mail. When requesting a priority gene, please indicate how variation analysis in this gene will enhance your NIEHS-funded research.

Home Gene Nomination Form Gene Targets A-Z Finished Genes Directory Finished Genes Table Proposed Genes Table Proposed Genes Table

Genotyping Resources Genotyping Methodology Genotyping Data Summary Genotyping Data Pathway Image Maps

## Haplotypes in NIEHS SNPs

National Institute of Environmental Health Sciences Environmental Genome Project NIEHS SNPs

Finished Genes (alphabetical directory by Hugo Gene Name)

A| B| C| D| E| F| G| H| I| K| L| M| N| O| P| Q| R| S| T| U| Z

MAP3L, MAP2A, MAP2B, MAP3K1, MAP3K2, MAP3K3, MAP3K4, MAP3K5, MAP3K6, MAP3K7, MAP3K8, MAP3K9, MAP3K10, MAP3K11, MAP3K12, MAP3K13, MAP3K14, MAP3K15, MAP3K16, MAP3K17, MAP3K18, MAP3K19, MAP3K20, MAP3K21, MAP3K22, MAP3K23, MAP3K24, MAP3K25, MAP3K26, MAP3K27, MAP3K28, MAP3K29, MAP3K30, MAP3K31, MAP3K32, MAP3K33, MAP3K34, MAP3K35, MAP3K36, MAP3K37, MAP3K38, MAP3K39, MAP3K40, MAP3K41, MAP3K42, MAP3K43, MAP3K44, MAP3K45, MAP3K46, MAP3K47, MAP3K48, MAP3K49, MAP3K50, MAP3K51, MAP3K52, MAP3K53, MAP3K54, MAP3K55, MAP3K56, MAP3K57, MAP3K58, MAP3K59, MAP3K60, MAP3K61, MAP3K62, MAP3K63, MAP3K64, MAP3K65, MAP3K66, MAP3K67, MAP3K68, MAP3K69, MAP3K70, MAP3K71, MAP3K72, MAP3K73, MAP3K74, MAP3K75, MAP3K76, MAP3K77, MAP3K78, MAP3K79, MAP3K80, MAP3K81, MAP3K82, MAP3K83, MAP3K84, MAP3K85, MAP3K86, MAP3K87, MAP3K88, MAP3K89, MAP3K90, MAP3K91, MAP3K92, MAP3K93, MAP3K94, MAP3K95, MAP3K96, MAP3K97, MAP3K98, MAP3K99, MAP3K100, MAP3K101, MAP3K102, MAP3K103, MAP3K104, MAP3K105, MAP3K106, MAP3K107, MAP3K108, MAP3K109, MAP3K110, MAP3K111, MAP3K112, MAP3K113, MAP3K114, MAP3K115, MAP3K116, MAP3K117, MAP3K118, MAP3K119, MAP3K120, MAP3K121, MAP3K122, MAP3K123, MAP3K124, MAP3K125, MAP3K126, MAP3K127, MAP3K128, MAP3K129, MAP3K130, MAP3K131, MAP3K132, MAP3K133, MAP3K134, MAP3K135, MAP3K136, MAP3K137, MAP3K138, MAP3K139, MAP3K140, MAP3K141, MAP3K142, MAP3K143, MAP3K144, MAP3K145, MAP3K146, MAP3K147, MAP3K148, MAP3K149, MAP3K150, MAP3K151, MAP3K152, MAP3K153, MAP3K154, MAP3K155, MAP3K156, MAP3K157, MAP3K158, MAP3K159, MAP3K160, MAP3K161, MAP3K162, MAP3K163, MAP3K164, MAP3K165, MAP3K166, MAP3K167, MAP3K168, MAP3K169, MAP3K170, MAP3K171, MAP3K172, MAP3K173, MAP3K174, MAP3K175, MAP3K176, MAP3K177, MAP3K178, MAP3K179, MAP3K180, MAP3K181, MAP3K182, MAP3K183, MAP3K184, MAP3K185, MAP3K186, MAP3K187, MAP3K188, MAP3K189, MAP3K190, MAP3K191, MAP3K192, MAP3K193, MAP3K194, MAP3K195, MAP3K196, MAP3K197, MAP3K198, MAP3K199, MAP3K200, MAP3K201, MAP3K202, MAP3K203, MAP3K204, MAP3K205, MAP3K206, MAP3K207, MAP3K208, MAP3K209, MAP3K210, MAP3K211, MAP3K212, MAP3K213, MAP3K214, MAP3K215, MAP3K216, MAP3K217, MAP3K218, MAP3K219, MAP3K220, MAP3K221, MAP3K222, MAP3K223, MAP3K224, MAP3K225, MAP3K226, MAP3K227, MAP3K228, MAP3K229, MAP3K230, MAP3K231, MAP3K232, MAP3K233, MAP3K234, MAP3K235, MAP3K236, MAP3K237, MAP3K238, MAP3K239, MAP3K240, MAP3K241, MAP3K242, MAP3K243, MAP3K244, MAP3K245, MAP3K246, MAP3K247, MAP3K248, MAP3K249, MAP3K250, MAP3K251, MAP3K252, MAP3K253, MAP3K254, MAP3K255, MAP3K256, MAP3K257, MAP3K258, MAP3K259, MAP3K260, MAP3K261, MAP3K262, MAP3K263, MAP3K264, MAP3K265, MAP3K266, MAP3K267, MAP3K268, MAP3K269, MAP3K270, MAP3K271, MAP3K272, MAP3K273, MAP3K274, MAP3K275, MAP3K276, MAP3K277, MAP3K278, MAP3K279, MAP3K280, MAP3K281, MAP3K282, MAP3K283, MAP3K284, MAP3K285, MAP3K286, MAP3K287, MAP3K288, MAP3K289, MAP3K290, MAP3K291, MAP3K292, MAP3K293, MAP3K294, MAP3K295, MAP3K296, MAP3K297, MAP3K298, MAP3K299, MAP3K300, MAP3K301, MAP3K302, MAP3K303, MAP3K304, MAP3K305, MAP3K306, MAP3K307, MAP3K308, MAP3K309, MAP3K310, MAP3K311, MAP3K312, MAP3K313, MAP3K314, MAP3K315, MAP3K316, MAP3K317, MAP3K318, MAP3K319, MAP3K320, MAP3K321, MAP3K322, MAP3K323, MAP3K324, MAP3K325, MAP3K326, MAP3K327, MAP3K328, MAP3K329, MAP3K330, MAP3K331, MAP3K332, MAP3K333, MAP3K334, MAP3K335, MAP3K336, MAP3K337, MAP3K338, MAP3K339, MAP3K340, MAP3K341, MAP3K342, MAP3K343, MAP3K344, MAP3K345, MAP3K346, MAP3K347, MAP3K348, MAP3K349, MAP3K350, MAP3K351, MAP3K352, MAP3K353, MAP3K354, MAP3K355, MAP3K356, MAP3K357, MAP3K358, MAP3K359, MAP3K360, MAP3K361, MAP3K362, MAP3K363, MAP3K364, MAP3K365, MAP3K366, MAP3K367, MAP3K368, MAP3K369, MAP3K370, MAP3K371, MAP3K372, MAP3K373, MAP3K374, MAP3K375, MAP3K376, MAP3K377, MAP3K378, MAP3K379, MAP3K380, MAP3K381, MAP3K382, MAP3K383, MAP3K384, MAP3K385, MAP3K386, MAP3K387, MAP3K388, MAP3K389, MAP3K390, MAP3K391, MAP3K392, MAP3K393, MAP3K394, MAP3K395, MAP3K396, MAP3K397, MAP3K398, MAP3K399, MAP3K400, MAP3K401, MAP3K402, MAP3K403, MAP3K404, MAP3K405, MAP3K406, MAP3K407, MAP3K408, MAP3K409, MAP3K410, MAP3K411, MAP3K412, MAP3K413, MAP3K414, MAP3K415, MAP3K416, MAP3K417, MAP3K418, MAP3K419, MAP3K420, MAP3K421, MAP3K422, MAP3K423, MAP3K424, MAP3K425, MAP3K426, MAP3K427, MAP3K428, MAP3K429, MAP3K430, MAP3K431, MAP3K432, MAP3K433, MAP3K434, MAP3K435, MAP3K436, MAP3K437, MAP3K438, MAP3K439, MAP3K440, MAP3K441, MAP3K442, MAP3K443, MAP3K444, MAP3K445, MAP3K446, MAP3K447, MAP3K448, MAP3K449, MAP3K450, MAP3K451, MAP3K452, MAP3K453, MAP3K454, MAP3K455, MAP3K456, MAP3K457, MAP3K458, MAP3K459, MAP3K460, MAP3K461, MAP3K462, MAP3K463, MAP3K464, MAP3K465, MAP3K466, MAP3K467, MAP3K468, MAP3K469, MAP3K470, MAP3K471, MAP3K472, MAP3K473, MAP3K474, MAP3K475, MAP3K476, MAP3K477, MAP3K478, MAP3K479, MAP3K480, MAP3K481, MAP3K482, MAP3K483, MAP3K484, MAP3K485, MAP3K486, MAP3K487, MAP3K488, MAP3K489, MAP3K490, MAP3K491, MAP3K492, MAP3K493, MAP3K494, MAP3K495, MAP3K496, MAP3K497, MAP3K498, MAP3K499, MAP3K500, MAP3K501, MAP3K502, MAP3K503, MAP3K504, MAP3K505, MAP3K506, MAP3K507, MAP3K508, MAP3K509, MAP3K510, MAP3K511, MAP3K512, MAP3K513, MAP3K514, MAP3K515, MAP3K516, MAP3K517, MAP3K518, MAP3K519, MAP3K520, MAP3K521, MAP3K522, MAP3K523, MAP3K524, MAP3K525, MAP3K526, MAP3K527, MAP3K528, MAP3K529, MAP3K530, MAP3K531, MAP3K532, MAP3K533, MAP3K534, MAP3K535, MAP3K536, MAP3K537, MAP3K538, MAP3K539, MAP3K540, MAP3K541, MAP3K542, MAP3K543, MAP3K544, MAP3K545, MAP3K546, MAP3K547, MAP3K548, MAP3K549, MAP3K550, MAP3K551, MAP3K552, MAP3K553, MAP3K554, MAP3K555, MAP3K556, MAP3K557, MAP3K558, MAP3K559, MAP3K560, MAP3K561, MAP3K562, MAP3K563, MAP3K564, MAP3K565, MAP3K566, MAP3K567, MAP3K568, MAP3K569, MAP3K570, MAP3K571, MAP3K572, MAP3K573, MAP3K574, MAP3K575, MAP3K576, MAP3K577, MAP3K578, MAP3K579, MAP3K580, MAP3K581, MAP3K582, MAP3K583, MAP3K584, MAP3K585, MAP3K586, MAP3K587, MAP3K588, MAP3K589, MAP3K590, MAP3K591, MAP3K592, MAP3K593, MAP3K594, MAP3K595, MAP3K596, MAP3K597, MAP3K598, MAP3K599, MAP3K600, MAP3K601, MAP3K602, MAP3K603, MAP3K604, MAP3K605, MAP3K606, MAP3K607, MAP3K608, MAP3K609, MAP3K610, MAP3K611, MAP3K612, MAP3K613, MAP3K614, MAP3K615, MAP3K616, MAP3K617, MAP3K618, MAP3K619, MAP3K620, MAP3K621, MAP3K622, MAP3K623, MAP3K624, MAP3K625, MAP3K626, MAP3K627, MAP3K628, MAP3K629, MAP3K630, MAP3K631, MAP3K632, MAP3K633, MAP3K634, MAP3K635, MAP3K636, MAP3K637, MAP3K638, MAP3K639, MAP3K640, MAP3K641, MAP3K642, MAP3K643, MAP3K644, MAP3K645, MAP3K646, MAP3K647, MAP3K648, MAP3K649, MAP3K650, MAP3K651, MAP3K652, MAP3K653, MAP3K654, MAP3K655, MAP3K656, MAP3K657, MAP3K658, MAP3K659, MAP3K660, MAP3K661, MAP3K662, MAP3K663, MAP3K664, MAP3K665, MAP3K666, MAP3K667, MAP3K668, MAP3K669, MAP3K670, MAP3K671, MAP3K672, MAP3K673, MAP3K674, MAP3K675, MAP3K676, MAP3K677, MAP3K678, MAP3K679, MAP3K680, MAP3K681, MAP3K682, MAP3K683, MAP3K684, MAP3K685, MAP3K686, MAP3K687, MAP3K688, MAP3K689, MAP3K690, MAP3K691, MAP3K692, MAP3K693, MAP3K694, MAP3K695, MAP3K696, MAP3K697, MAP3K698, MAP3K699, MAP3K700, MAP3K701, MAP3K702, MAP3K703, MAP3K704, MAP3K705, MAP3K706, MAP3K707, MAP3K708, MAP3K709, MAP3K710, MAP3K711, MAP3K712, MAP3K713, MAP3K714, MAP3K715, MAP3K716, MAP3K717, MAP3K718, MAP3K719, MAP3K720, MAP3K721, MAP3K722, MAP3K723, MAP3K724, MAP3K725, MAP3K726, MAP3K727, MAP3K728, MAP3K729, MAP3K730, MAP3K731, MAP3K732, MAP3K733, MAP3K734, MAP3K735, MAP3K736, MAP3K737, MAP3K738, MAP3K739, MAP3K740, MAP3K741, MAP3K742, MAP3K743, MAP3K744, MAP3K745, MAP3K746, MAP3K747, MAP3K748, MAP3K749, MAP3K750, MAP3K751, MAP3K752, MAP3K753, MAP3K754, MAP3K755, MAP3K756, MAP3K757, MAP3K758, MAP3K759, MAP3K760, MAP3K761, MAP3K762, MAP3K763, MAP3K764, MAP3K765, MAP3K766, MAP3K767, MAP3K768, MAP3K769, MAP3K770, MAP3K771, MAP3K772, MAP3K773, MAP3K774, MAP3K775, MAP3K776, MAP3K777, MAP3K778, MAP3K779, MAP3K780, MAP3K781, MAP3K782, MAP3K783, MAP3K784, MAP3K785, MAP3K786, MAP3K787, MAP3K788, MAP3K789, MAP3K790, MAP3K791, MAP3K792, MAP3K793, MAP3K794, MAP3K795, MAP3K796, MAP3K797, MAP3K798, MAP3K799, MAP3K800, MAP3K801, MAP3K802, MAP3K803, MAP3K804, MAP3K805, MAP3K806, MAP3K807, MAP3K808, MAP3K809, MAP3K810, MAP3K811, MAP3K812, MAP3K813, MAP3K814, MAP3K815, MAP3K816, MAP3K817, MAP3K818, MAP3K819, MAP3K820, MAP3K821, MAP3K822, MAP3K823, MAP3K824, MAP3K825, MAP3K826, MAP3K827, MAP3K828, MAP3K829, MAP3K830, MAP3K831, MAP3K832, MAP3K833, MAP3K834, MAP3K835, MAP3K836, MAP3K837, MAP3K838, MAP3K839, MAP3K840, MAP3K841, MAP3K842, MAP3K843, MAP3K844, MAP3K845, MAP3K846, MAP3K847, MAP3K848, MAP3K849, MAP3K850, MAP3K851, MAP3K852, MAP3K853, MAP3K854, MAP3K855, MAP3K856, MAP3K857, MAP3K858, MAP3K859, MAP3K860, MAP3K861, MAP3K862, MAP3K863, MAP3K864, MAP3K865, MAP3K866, MAP3K867, MAP3K868, MAP3K869, MAP3K870, MAP3K871, MAP3K872, MAP3K873, MAP3K874, MAP3K875, MAP3K876, MAP3K877, MAP3K878, MAP3K879, MAP3K880, MAP3K881, MAP3K882, MAP3K883, MAP3K884, MAP3K885, MAP3K886, MAP3K887, MAP3K888, MAP3K889, MAP3K890, MAP3K891, MAP3K892, MAP3K893, MAP3K894, MAP3K895, MAP3K896, MAP3K897, MAP3K898, MAP3K899, MAP3K900, MAP3K901, MAP3K902, MAP3K903, MAP3K904, MAP3K905, MAP3K906, MAP3K907, MAP3K908, MAP3K909, MAP3K910, MAP3K911, MAP3K912, MAP3K913, MAP3K914, MAP3K915, MAP3K916, MAP3K917, MAP3K918, MAP3K919, MAP3K920, MAP3K921, MAP3K922, MAP3K923, MAP3K924, MAP3K925, MAP3K926, MAP3K927, MAP3K928, MAP3K929, MAP3K930, MAP3K931, MAP3K932, MAP3K933, MAP3K934, MAP3K935, MAP3K936, MAP3K937, MAP3K938, MAP3K939, MAP3K940, MAP3K941, MAP3K942, MAP3K943, MAP3K944, MAP3K945, MAP3K946, MAP3K947, MAP3K948, MAP3K949, MAP3K950, MAP3K951, MAP3K952, MAP3K953, MAP3K954, MAP3K955, MAP3K956, MAP3K957, MAP3K958, MAP3K959, MAP3K960, MAP3K961, MAP3K962, MAP3K963, MAP3K964, MAP3K965, MAP3K966, MAP3K967, MAP3K968, MAP3K969, MAP3K970, MAP3K971, MAP3K972, MAP3K973, MAP3K974, MAP3K975, MAP3K976, MAP3K977, MAP3K978, MAP3K979, MAP3K980, MAP3K981, MAP3K982, MAP3K983, MAP3K984, MAP3K985, MAP3K986, MAP3K987, MAP3K988, MAP3K989, MAP3K990, MAP3K991, MAP3K992, MAP3K993, MAP3K994, MAP3K995, MAP3K996, MAP3K997, MAP3K998, MAP3K999, MAP3K1000, MAP3K1001, MAP3K1002, MAP3K1003, MAP3K1004, MAP3K1005, MAP3K1006, MAP3K1007, MAP3K1008, MAP3K1009, MAP3K1010, MAP3K1011, MAP3K1012, MAP3K1013, MAP3K1014, MAP3K1015, MAP3K1016, MAP3K1017, MAP3K1018, MAP3K1019, MAP3K1020, MAP3K1021, MAP3K1022, MAP3K1023, MAP3K1024, MAP3K1025, MAP3K1026, MAP3K1027, MAP3K1028, MAP3K1029, MAP3K1030, MAP3K1031, MAP3K1032, MAP3K1033, MAP3K1034, MAP3K1035, MAP3K1036, MAP3K1037, MAP3K1038, MAP3K1039, MAP3K1040, MAP3K1041, MAP3K1042, MAP3K1043, MAP3K1044, MAP3K1045, MAP3K1046, MAP3K1047, MAP3K1048, MAP3K1049, MAP3K1050, MAP3K1051, MAP3K1052, MAP3K1053, MAP3K1054, MAP3K1055, MAP3K1056, MAP3K1057, MAP3K1058, MAP3K1059, MAP3K1060, MAP3K1061, MAP3K1062, MAP3K1063, MAP3K1064, MAP3K1065, MAP3K1066, MAP3K1067, MAP3K1068, MAP3K1069, MAP3K1070, MAP3K1071, MAP3K1072, MAP3K1073, MAP3K1074, MAP3K1075, MAP3K1076, MAP3K1077, MAP3K1078, MAP3K1079, MAP3K1080, MAP3K1081, MAP3K1082, MAP3K1083, MAP3K1084, MAP3K1085, MAP3K1086, MAP3K1087, MAP3K1088, MAP3K1089, MAP3K1090, MAP3K1091, MAP3K1092, MAP3K1093, MAP3K1094, MAP3K1095, MAP3K1096, MAP3K1097, MAP3K1098, MAP3K1099, MAP3K1100, MAP3K1101, MAP3K1102, MAP3K1103, MAP3K1104, MAP3K1105, MAP3K1106, MAP3K1107, MAP3K1108, MAP3K1109, MAP3K1110, MAP3K1111, MAP3K1112, MAP3K1113, MAP3K1114, MAP3K1115, MAP3K1116, MAP3K1117, MAP3K1118, MAP3K1119, MAP3K1120, MAP3K1121, MAP3K1122, MAP3K1123, MAP3K1124, MAP3K1125, MAP3K1126, MAP3K1127, MAP3K1128, MAP3K1129, MAP3K1130, MAP3K1131, MAP3K1132, MAP3K1133, MAP3K1134, MAP3K1135, MAP3K1136, MAP3K1137, MAP3K1138, MAP3K1139, MAP3K1140, MAP3K1141, MAP3K1142, MAP3K1143, MAP3K1144, MAP3K1145, MAP3K1146, MAP3K1147, MAP3K1148, MAP3K1149, MAP3K1150, MAP3K1151, MAP3K1152, MAP3K1153, MAP3K1154, MAP3K1155, MAP3K1156, MAP3K1157, MAP3K1158, MAP3K1159, MAP3K1160, MAP3K1161, MAP3K1162, MAP3K1163, MAP3K1164, MAP3K1165, MAP3K1166, MAP3K1167, MAP3K1168, MAP3K1169, MAP3K1170, MAP3K1171, MAP3K1172, MAP3K1173, MAP3K1174, MAP3K1175, MAP3K1176, MAP3K1177, MAP3K1178, MAP3K1179, MAP3K1180, MAP3K1181, MAP3K1182, MAP3K1183, MAP3K1184, MAP3K1185, MAP3K1186, MAP3K1187, MAP3K1188, MAP3K1189, MAP3K1190, MAP3K1191, MAP3K1192, MAP3K1193, MAP3K1194, MAP3K1195, MAP3K1196, MAP3K1197, MAP3K1198, MAP3K1199, MAP3K1200, MAP3K1201, MAP3K1202, MAP3K1203, MAP3K1204, MAP3K1205, MAP3K1206, MAP3K1207, MAP3K1208, MAP3K1209, MAP3K1210, MAP3K1211, MAP3K1212, MAP3K1213, MAP3K1214, MAP3K1215, MAP3K1216, MAP3K1217, MAP3K1218, MAP3K1219, MAP3K1220, MAP3K1221, MAP3K1222, MAP3K1223, MAP3K1224, MAP3K1225, MAP3K1226, MAP3K1227, MAP3K1228, MAP3K1229, MAP3K1230, MAP3K1231, MAP3K1232, MAP3K1233, MAP3K1234, MAP3K1235, MAP3K1236, MAP3K1237, MAP3K1238, MAP3K1239, MAP3K1240, MAP3K1241, MAP3K1242, MAP3K1243, MAP3K1244, MAP3K1245, MAP3K1246, MAP3K1247, MAP3K1248, MAP3K1249, MAP3K1250, MAP3K1251, MAP3K1252, MAP3K1253, MAP3K1254, MAP3K1255, MAP3K1256, MAP3K1257, MAP3K1258, MAP3K1259, MAP3K1260, MAP3K1261, MAP3K1262, MAP3K1263, MAP3K1264, MAP3K1265, MAP3K1266, MAP3K1267, MAP3K1268, MAP3K1269, MAP3K1270, MAP3K1271, MAP3K1272, MAP3K1273, MAP3K1274, MAP3K1275, MAP3K1276, MAP3K1277, MAP3K1278, MAP3K1279, MAP3K1280, MAP3K1281, MAP3K1282, MAP3K1283, MAP3K1284, MAP3K1285, MAP3K1286, MAP3K1287, MAP3K1288, MAP3K1289, MAP3K1290, MAP3K1291, MAP3K1292, MAP3K1293, MAP3K1294, MAP3K1295, MAP3K1296, MAP3K1297, MAP3K1298, MAP3K1299, MAP3K1300, MAP3K1301, MAP3K1302, MAP3K1303, MAP3K1304, MAP3K1305, MAP3K1306, MAP3K1307, MAP3K1308, MAP3K1309, MAP3K1310, MAP3K1311, MAP3K1312, MAP3K1313, MAP3K1314, MAP3K1315, MAP3K1316, MAP3K1317, MAP3K1318, MAP3K1319, MAP3K1320, MAP3K1321, MAP3K1322, MAP3K1323, MAP3K1324, MAP3K1325, MAP3K1326, MAP3K1327, MAP3K1328, MAP3K1329, MAP3K1330, MAP3K1331, MAP3K1332, MAP3K1333, MAP3K1334, MAP3K1335, MAP3K1336, MAP3K1337, MAP3K1338, MAP3K1339, MAP3K1340, MAP3K1341, MAP3K1342, MAP3K1343, MAP3K1344, MAP3K1345, MAP3K1346, MAP3K1347, MAP3K1348, MAP3K1349, MAP3K1350, MAP3K1351, MAP3K1352, MAP3K1353, MAP3K1354, MAP3K1355, MAP3K1356, MAP3K1357, MAP3K1358, MAP3K1359, MAP3K1360, MAP3K1361, MAP3K1362, MAP3K1363, MAP3K1364, MAP3K1365, MAP3K1366, MAP3K1367, MAP3K1368, MAP3K1369, MAP3K1370, MAP3K1371, MAP3K1372, MAP3K1373, MAP3K1374, MAP3K1375, MAP3K1376, MAP3K1377, MAP3K1378, MAP3K1379, MAP3K1380, MAP3K1381, MAP3K1382, MAP3K1383, MAP3K1384, MAP3K1385, MAP3K1386, MAP3K1387, MAP3K1388, MAP3K1389, MAP3K1390, MAP3K1391, MAP3K1392, MAP3K1393, MAP3K1394, MAP3K

# Haplotypes in NIEHS SNPs

Return to NIEHS SNP's Home  
Return to A21 Selected Genes Directory

**ME:myoglobin**  
Chromosomal Location: 22q13.1

National Institute of Environmental Health Sciences  
National Institutes of Health  
NIH

GeneSNP Image

View	Full Gene	SNP Type	SNP ID	Repeat	SNP ID	Frequency	Reset
	1000 bp	SNP	SNP	None	All SNPs	High Freq	

Gene-Specific Links

Entrez Gene      Golden Path (NCSC Genome Browser)      Golden Path (with NIEHS SNP's Tracks)      Pub Med

Downloaded 29 files of all data for this gene      Sample Population Description

	SNPs cDNA	Color FASTA SNP Called	PCR Primers (FASTA) Genbank
	Visual Genotype Individual Genotypes	SNP Alleles SNP Allele Frequency	SNP-Hardy-Weinberg

The figure shows a screenshot of a web-based bioinformatics tool. At the top, a large yellow header reads "Haplotypes in NIEHS SNPs". Below the header is a grid of SNP data. The left side of the grid has columns for "SNP ID", "Chromosome", "Position", "Allele", "Genotype", and "P-value". The right side has columns for "Haplotype", "Frequency", and "P-value". Above the grid, there are two sections: "IMAGE CONTROLS" and "SNP CONTROLS". Each section has dropdown menus for "View" (Full Gene or 1000 bp), "Search by" (SNP ID or Position), and "Repeats" (None, Low, High). Below these sections are buttons for "Best" and "Reset".

**Gene-Specific Links**

Entrez Gene    GoPubMed (NCBI Genome Browser)    GoldenPath (with NIEHS SNPs Tracks)    Pub Med

Download a zip file of all data for this gene    Sample Population Description

SNP Data	Mapping Data	cSNPs cDNA	Color FASTA SNP Conted	PCR Primers (FASTA) Genbank	
Genotyping Data		Visual Genotype Individual Genotypes	SNP Alleles SNP Allele Frequency	SNP-Hardy-Weinberg	
Haplotyping Data		PHASE Output	Phased Individual Haplotypes	Sorted by Frequency	
Linkage Data		Visual Haplotype			
Predictive Analyses		LD Select (Tag SNPs) African Descent	European Descent	Hispanic Descent	Asian Descent
		Nonredundant cSNP Analysis			

The figure shows a screenshot of the Haplotype viewer interface. At the top, a large yellow header reads "Haplotypes in NIEHS SNPs". Below the header is a navigation bar with tabs: "SNP", "Haplotype", "Marker", "SNP+Haplotype", and "SNP+Marker". The main content area features a grid of SNPs with various alleles. A legend at the top right indicates SNP types: SNPs (green), Haplotype SNPs (blue), and Marker SNPs (red). Below the grid are two search boxes: "IMAGE CONTROLS" (View: FullGene, Search: 100 bp) and "SNP CONTROLS" (View: All SNPs, Frequency: High). A "Reset" button is located at the bottom right of each search box. A "Gene-Specific Links" section follows, containing four buttons: "Genome Path (NCBI Genome Browser)", "Genome Path (with NIEHS SNPs Tracks)", "Pub Med", and "Population Description". Below this is a "Download a zip file of all data for this gene" link. The interface then displays five data categories in rows:

- Mapping Data**: Shows a small map icon and a "View" button.
- Gendotyping Data**: Shows a 4x4 grid icon and a "View" button.
- Haplotyping Data**: Shows a 4x4 grid icon and a "View" button.
- Linkage Data**: Shows a pedigree icon and a "View" button.
- Predictive Analyses**: Shows a DNA helix icon and a "View" button.

Each row contains three columns of data corresponding to the "View" button, such as "cSNPs", "cDNA", "Color FASTA", "SNP Contect", "SNP Alleles", "SNP Hardy-Weinberg", "PhASE Output", "Haplotype", and "Sorted by Frequency". The "Haplotype" column for the Haplotyping Data row is highlighted with a green border.

 <h1>Haplotypes in NIEHS SNPs</h1>																																									
<p>National Institute of Environmental Health Sciences Environmental Genome Project</p>																																									
<h2>NIEHS SNPs</h2>																																									
<p><b>Name</b></p> <p><b>Gene Nominations Form</b></p> <p><b>Gene Targets</b></p> <ul style="list-style-type: none"> <li>• <a href="#">Entrez Gene</a></li> <li>• <a href="#">Ensembl Gene</a></li> <li>• <a href="#">UniProt</a></li> <li>• <a href="#">Protein Data Bank</a></li> <li>• <a href="#">Primate Gene Database</a></li> </ul> <p><b>Genotyping Resources</b></p> <ul style="list-style-type: none"> <li>• <a href="#">SNPedia</a></li> <li>• <a href="#">dbSNP</a></li> <li>• <a href="#">SNPedia Selected</a></li> <li>• <a href="#">SNPedia Selected</a></li> <li>• <a href="#">SNPedia Selected</a></li> </ul> <p><b>Pathway Enrichment Data</b></p> <ul style="list-style-type: none"> <li>• <a href="#">KEGG Pathway</a></li> <li>• <a href="#">Reactome</a></li> <li>• <a href="#">BIOCARTA</a></li> <li>• <a href="#">Double Strand Break Repair</a></li> <li>• <a href="#">MicroRNA Target</a></li> <li>• <a href="#">Macrophage Cytokine Response</a></li> </ul> <p><b>Personal</b></p> <p><b>Summary Information</b></p> <ul style="list-style-type: none"> <li>• <a href="#">Curated Genes</a></li> <li>• <a href="#">Curated Genes</a></li> <li>• <a href="#">Data Downloads</a></li> </ul> <p><b>Design Policy</b></p> <p><b>Visual Genotypes</b></p> <p><b>Visual Genotypes</b> </p>	<p><b>Last Update:</b> <a href="#">Full Dataset</a></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Dataset</th> <th>Last Update</th> </tr> </thead> <tbody> <tr> <td>SNPedia</td> <td>December 20, 2005</td> </tr> <tr> <td>FANCO</td> <td>December 19, 2005</td> </tr> <tr> <td>MARCKS</td> <td>December 19, 2005</td> </tr> <tr> <td>ALKBHD</td> <td>December 12, 2005</td> </tr> <tr> <td>ANXA10</td> <td>December 10, 2005</td> </tr> <tr> <td>ATM</td> <td>November 30, 2005</td> </tr> <tr> <td>CD36</td> <td>November 30, 2005</td> </tr> <tr> <td>DGAT1</td> <td>November 30, 2005</td> </tr> <tr> <td>GPR158</td> <td>November 30, 2005</td> </tr> <tr> <td>MEF2C</td> <td>November 30, 2005</td> </tr> <tr> <td>MRN</td> <td>November 30, 2005</td> </tr> <tr> <td>DRD2</td> <td>November 30, 2005</td> </tr> <tr> <td>DRD5</td> <td>November 30, 2005</td> </tr> <tr> <td>CYCB6</td> <td>November 22, 2005</td> </tr> <tr> <td>MTG8</td> <td>November 22, 2005</td> </tr> <tr> <td>PREST</td> <td>November 22, 2005</td> </tr> <tr> <td>PTEN</td> <td>November 22, 2005</td> </tr> <tr> <td>MATN1</td> <td>November 1, 2005</td> </tr> <tr> <td>SER1</td> <td>November 1, 2005</td> </tr> </tbody> </table>	Dataset	Last Update	SNPedia	December 20, 2005	FANCO	December 19, 2005	MARCKS	December 19, 2005	ALKBHD	December 12, 2005	ANXA10	December 10, 2005	ATM	November 30, 2005	CD36	November 30, 2005	DGAT1	November 30, 2005	GPR158	November 30, 2005	MEF2C	November 30, 2005	MRN	November 30, 2005	DRD2	November 30, 2005	DRD5	November 30, 2005	CYCB6	November 22, 2005	MTG8	November 22, 2005	PREST	November 22, 2005	PTEN	November 22, 2005	MATN1	November 1, 2005	SER1	November 1, 2005
Dataset	Last Update																																								
SNPedia	December 20, 2005																																								
FANCO	December 19, 2005																																								
MARCKS	December 19, 2005																																								
ALKBHD	December 12, 2005																																								
ANXA10	December 10, 2005																																								
ATM	November 30, 2005																																								
CD36	November 30, 2005																																								
DGAT1	November 30, 2005																																								
GPR158	November 30, 2005																																								
MEF2C	November 30, 2005																																								
MRN	November 30, 2005																																								
DRD2	November 30, 2005																																								
DRD5	November 30, 2005																																								
CYCB6	November 22, 2005																																								
MTG8	November 22, 2005																																								
PREST	November 22, 2005																																								
PTEN	November 22, 2005																																								
MATN1	November 1, 2005																																								
SER1	November 1, 2005																																								
<p>Welcome to the NIEHS SNPs Program</p>																																									
<h3>Introduction</h3> <p>The NIEHS Environmental Genome Project is a multi-disciplinary, collaborative effort focused on examining the genetic basis for environmental exposures, interindividual response variation in human genes, and disease risk in U.S. populations. The NIEHS SNP Program at the University of Washington is targeted on identifying common sequence variation in human genes that may contribute to individual differences in response genes. The first phase of this effort is focused on finding common sequence variation (SNPs) in human genes involved in DNA repair and cell cycle pathways (see links on Gene Targets in the sidebar). The methods and concepts developed in this program can be applied to evaluating human disease risk with environmental exposures.</p>																																									
<h3>NIEHS SNPs Database</h3> <p>NIEHS SNPs are available in the <a href="#">GeneSNP</a> database as well as the national database resource, <a href="#">dbSNP</a>. GeneSNP provides a gene-centric view of the genome structure, coding sequences, and polymorphisms. dbSNP is a public database of human SNPs and includes a graphical view of all available SNP data including allele frequencies and genotypes in select populations. A key to understanding the polymorphic sites needs to examine disease risk in human population studies.</p>																																									
<h3>Polymorphism Analysis</h3> <p>Automated DNA sequencing is used to identify and generate SNPs in human candidate genes (see <a href="#">Protocol</a>). Candidate genes are being sequenced to identify common sequence variation for functional analysis and population-based studies. Candidate genes were formally sequenced across a panel of 90 individuals from diverse ethnic backgrounds. The results of this study have been published (<a href="#">Kruglyak et al., 2002</a>). SNPs are now being sequenced across a panel of 95 individuals of known ethnicities (see <a href="#">Sample Population</a>).</p>																																									

# Haplotypes in NIEHS SNPs

National Institute of Environmental Health Sciences  
Environmental Genome Project  
**NIEHS SNPs**

Search Site

**Home**

**Gene Nominations Form**

**Gene Targets**  
Add New Target  
Delete Target  
Update Gene Targets  
Proposed Gene Table

**Genotyping Resources**  
Download Genotype Data  
Download Genotype Data  
Download Genotype Data  
Download Genotype Data

**Pathway Image Maps**  
Apolipoprotein B  
C6  
Cdk5  
Dendrite  
Dendrite  
Dendrite  
Endothelial Enzyme  
Endothelial Enzyme

**Personalized**

**Summary Information for Competed Genes**  
Summary Statistics  
Download Genotype Data

**Usage Policy**

**Visual Genotypes**

**Visual Haplotypes**

Local Phastaceus Input File   
Phastaceus Score  Phastaceus Sample

EUP-Finished-Dna-Phastaceus.txt File  [NB]

Raw Allele Percentage Integer  MAUDOL2  
MAPIV24  
MAPV41

Cluster and/or Draw Tree(s)  NOTHING  
Generate Representative Phastaceus  MAPV9  
Run Vrt on the Web  MED4

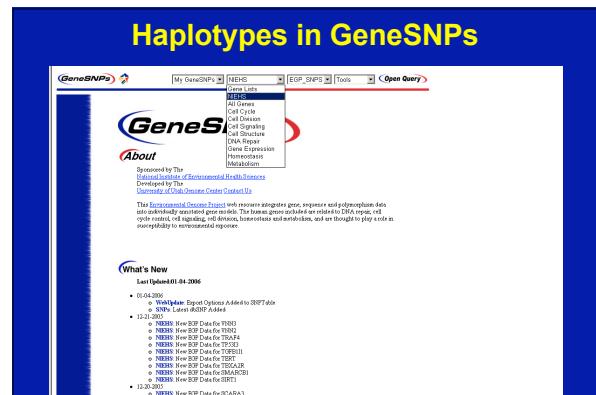
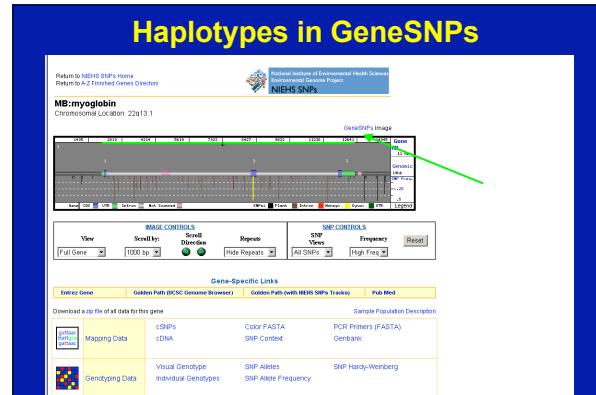
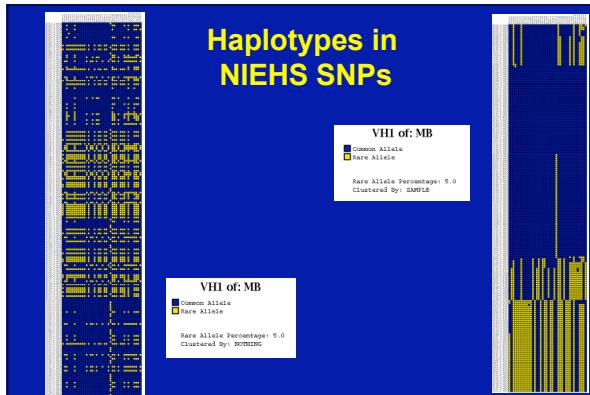
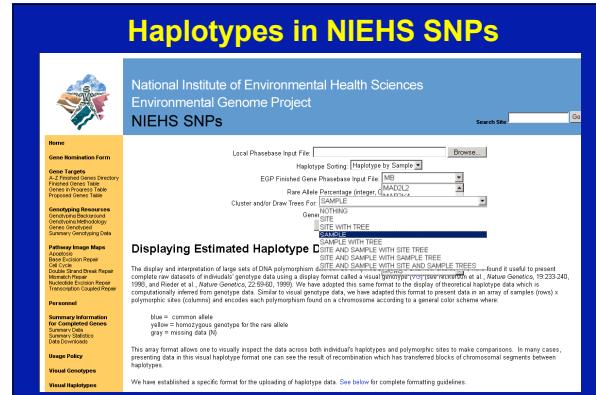
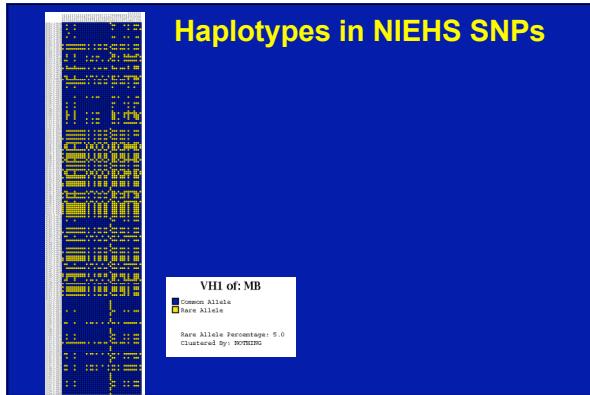
Phastaceus Input File  [MC1]  
[MC2]  
[MC3]  
[MC4]  
[MC5]

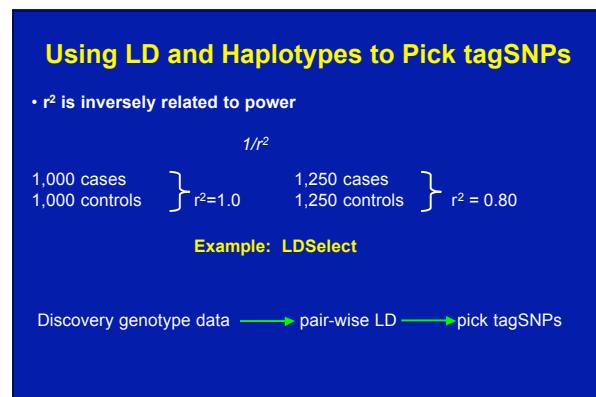
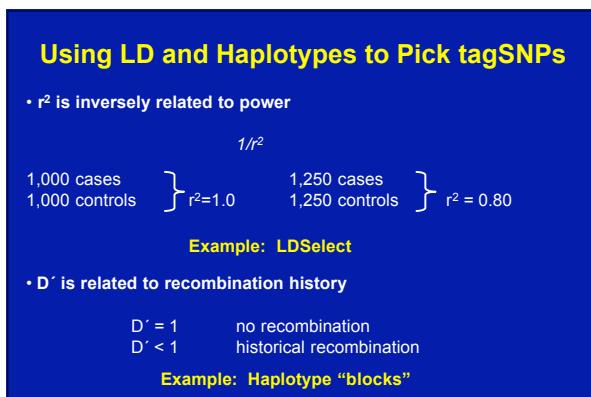
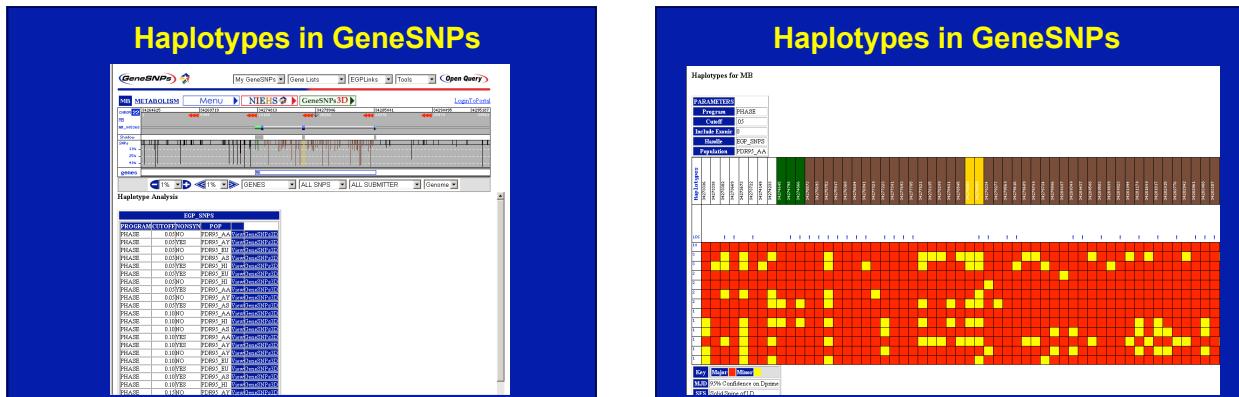
The display and interpretation of large sets of DNA polymorphism data can be simplified considerably by the use of haplotypes. A haplotype is a set of alleles at a number of loci (e.g., SNPs) that are transmitted together as a unit. The term haplotype refers to a particular combination of alleles at a number of loci. A haplotype can be thought of as a "genetic signature". Haplotype analysis is a powerful technique for understanding the genetic architecture of complex diseases. Similar to visual genotype data, we have adapted this format to present haplotype data in an array of samples (rows) x genomic sites (columns). We require each haplotype data as a chromosome according to a general convention where:

Blue = common allele  
Yellow = homozygous genotype to the rare allele

This array format allows one to quickly inspect the data across both individuals' haplotypes and polymorphic sites to make comparisons. In many cases, presenting data in this visual haplotype format can see the result of recombination which has been blocked in chromosomal segments between haplotypes.

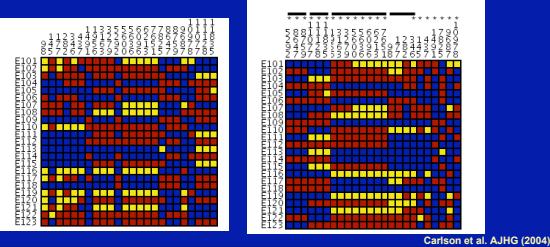
We have established a specific format for the uploading of haplotype data. See below for complete formatting guidelines.





## LDSelect: Using LD to Pick tagSNPs

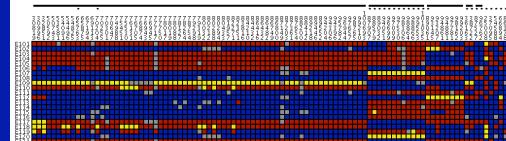
- LDSelect
    - Uses SNP discovery data (not haplotypes)
    - Finds all correlated SNPs to minimize the total number
    - Maintains genetic diversity of locus



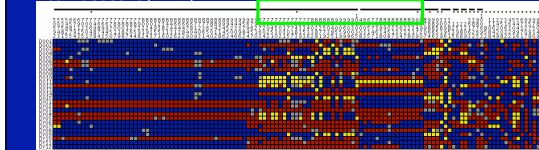
Carlson et al. AJHG (2004)

## TagSNPs Are Population Specific

## European-descent (BLM)



### African-descent (BLM)



## SNP Selection: tagSNP Data

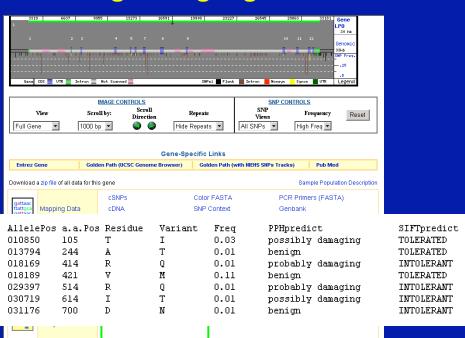
BLM

## Side Note: Categorizing taqSNPs

- SNP context  
Nonrepetitive > repetitive
  - Location of SNP  
Coding > noncoding
  - Function  
Nonsynonymous > synonymous

## Categorizing tagSNPs

LPO

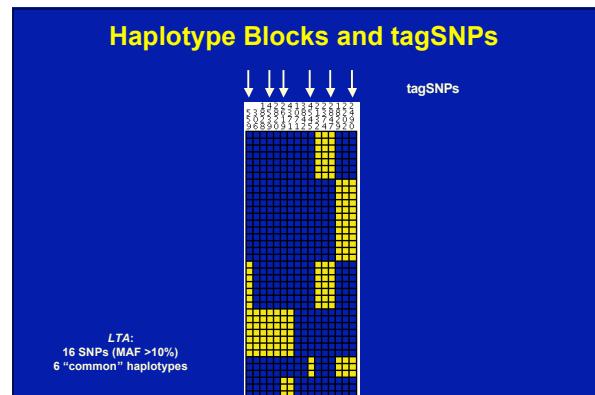
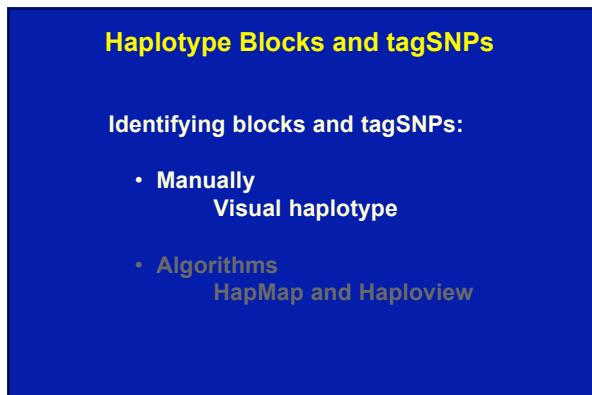
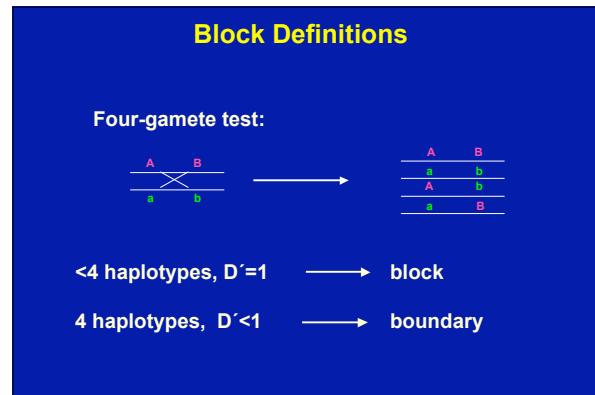
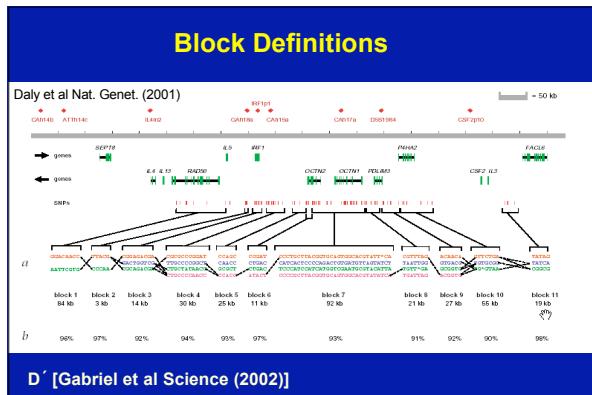
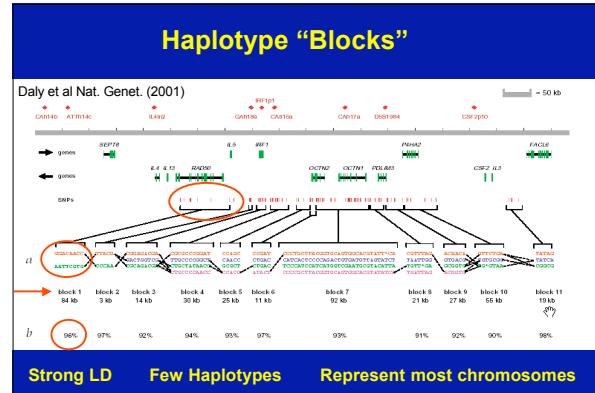
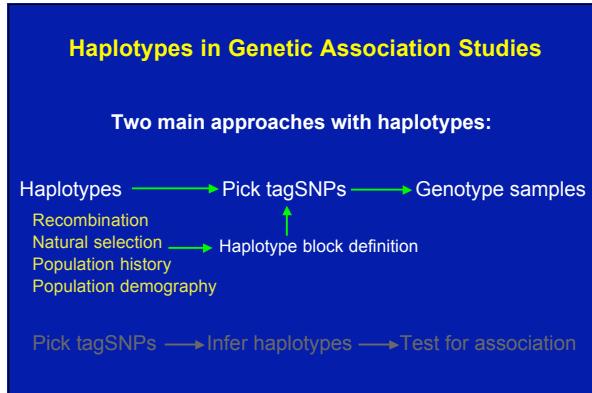


Haplotypes in Genetic Association Studies

## Two main approaches with haplotypes:

Haplotypes → Pick tagSNPs → Genotype samples

Pick tagSNPs → Infer haplotypes → Test for association



## Haplotype Blocks and tagSNPs

**Identifying blocks and tagSNPs:**

- Manually  
Visual Haplotype
- Algorithms  
HapMap and HaploView

## HapMap Data and Haplovview

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "About the International HapMap Project" for more information.

**Project Information**

- About the Project
- Haplotype Data
- Haplotype Publications
- Haplotype Tutorial
- Haplotype Mailing List
- Haplotype Project Participants
- Haplotype Minor Site in Japan

**Project Data**

**2005-10-24: HapMap Public Release #19**  
Genotypes, Sequences and assays for phase I and phase II of the HapMap project are now available for bulk download. The files contain all phase I and II data combined.

Populations	CEU	CHB	JPT	YRI
Total QC SNPs	3,967,409	3,803,524	3,802,523	3,809,307
Total QC SNPs with phase	3,967,409	3,803,524	3,802,523	3,809,307

**2005-05-26: Recombination rates and haplotypes for HapMap Phase II data released**  
Recombination rates and haplotypes are now available for bulk download.

**2005-05-26: HapMap Public Release #18**  
Genotypes, Sequences and assays for the chromosomes 2,6,11,14,15 and 21 are now available for bulk download only. Please note that this represents complete phase II data on these chromosomes for CEU and YRI samples. This is a dump with all phase I and II data combined. Genotype summary for CHB in this release is as follows:

Populations	CEU	CHB	JPT	YRI
Total QC SNPs	385,351	314,958	314,729	327,707
Chr 6	385,351	124,729	124,293	386,208
Chr 11	385,351	52,445	52,293	263,444
Chr 14	385,351	49,386	49,333	318,533
Chr 15	385,351	41,999	41,869	147,373
Chr 21	385,351	37,212	37,202	85,373
Total	1,217,105	870,515	870,515	1,610,773

[www.hapmap.org](http://www.hapmap.org)

**International HapMap Project**

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character '\*' is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

**Examples:** Chr20, Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Search**

**Reports & Analysis**

**Landmark or Region:** Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Data Source:** HapMap Data RefSeq1c phased June05, on NCBI B34 assembly, dbSNP b124

**Population descriptors:** YRI: Yoruba in Ibadan, Nigeria, JPT: Japanese in Tokyo, Japan, CHB: Han Chinese in Beijing, China, CEPH: Utah residents with ancestry from northern and western Europe

**Overview**

**Details**

**Tracks**

**Display Settings**

**Add your own tracks**

**International HapMap Project**

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character '\*' is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

**Examples:** Chr20, Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Search**

**Reports & Analysis**

**Landmark or Region:** Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Data Source:** HapMap Data RefSeq1c phased June05, on NCBI B34 assembly, dbSNP b124

**Population descriptors:** YRI: Yoruba in Ibadan, Nigeria, JPT: Japanese in Tokyo, Japan, CHB: Han Chinese in Beijing, China, CEPH: Utah residents with ancestry from northern and western Europe

**Overview**

**Details**

**Tracks**

**Display Settings**

**Add your own tracks**

**International HapMap Project**

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character '\*' is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

**Examples:** Chr20, Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Search**

**Reports & Analysis**

**Landmark or Region:** Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Data Source:** HapMap Data RefSeq1c phased June05, on NCBI B34 assembly, dbSNP b124

**Population descriptors:** YRI: Yoruba in Ibadan, Nigeria, JPT: Japanese in Tokyo, Japan, CHB: Han Chinese in Beijing, China, CEPH: Utah residents with ancestry from northern and western Europe

**Overview**

**Details**

**Tracks**

**Display Settings**

**Add your own tracks**

**International HapMap Project**

Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character '\*' is allowed. To center on a location, click the ruler. Use the ScrollZoom buttons to change magnification and position.

**Examples:** Chr20, Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Search**

**Reports & Analysis**

**Landmark or Region:** Chr6@60000,700000, SNP#6870669, NM\_153254, BRCA2, Sq31, Hapl10

**Data Source:** HapMap Data RefSeq1c phased June05, on NCBI B34 assembly, dbSNP b124

**Population descriptors:** YRI: Yoruba in Ibadan, Nigeria, JPT: Japanese in Tokyo, Japan, CHB: Han Chinese in Beijing, China, CEPH: Utah residents with ancestry from northern and western Europe

**Overview**

**Details**

**Tracks**

**Display Settings**

**Add your own tracks**

The screenshot shows the International HapMap Project website. A search bar at the top is set to "Showing 19.59 kbp from chr6, positions 133,045,579 to 133,065,167". Below it, a legend indicates "Alleles: Search using a sequence name, gene name, locus, or other landmark. The wildcard character \* is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position." A "Results" section displays a map of Europe and a detailed genomic track for chromosome 6. The track includes an ideogram, MT contigs, and a SNP density plot. A red arrow points to the "Population description" section, which states "Population description c790: Yoruba in Ibadan, Nigeria; JP1: Japanese in Tokyo, Japan; CHB: Han Chinese in Beijing, China; CEU: Utah residents with ancestry from northern and western Europe".

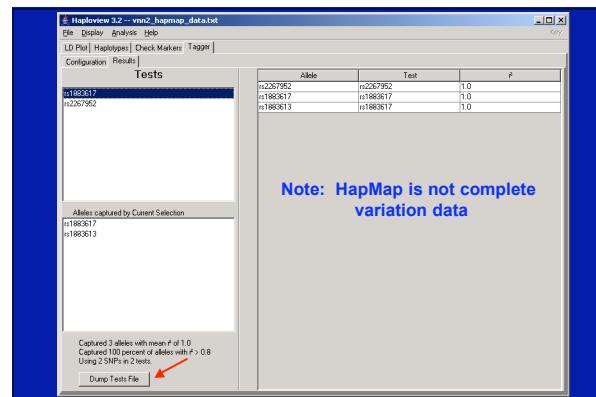
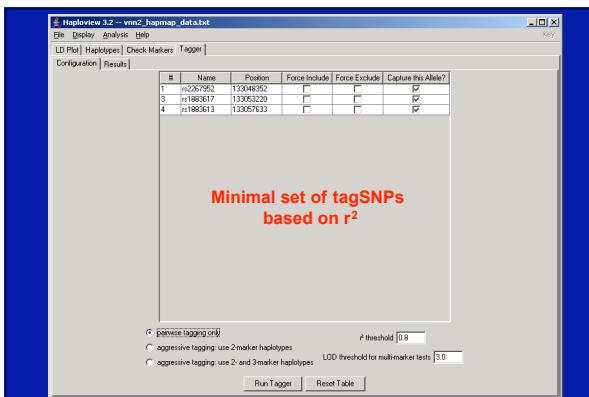
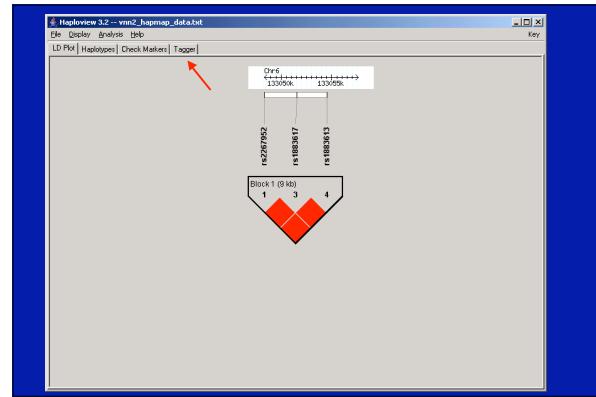
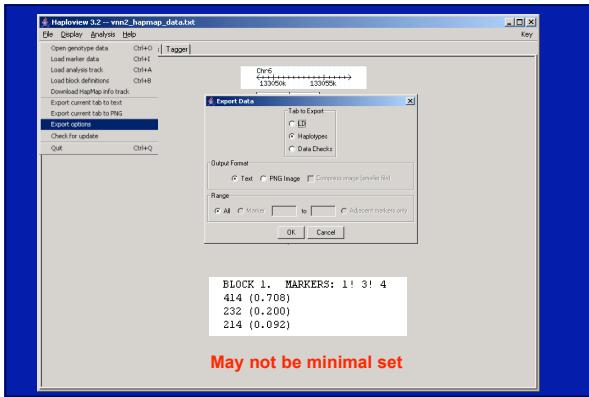
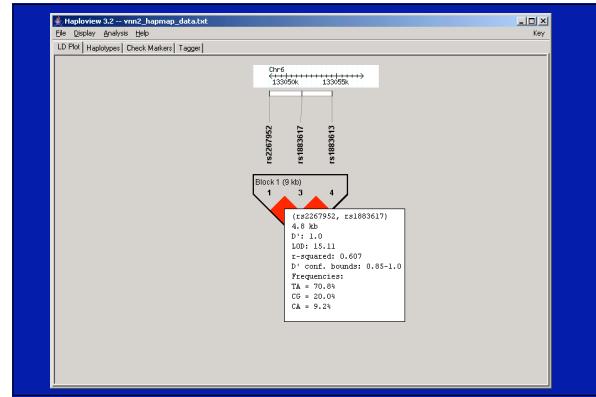
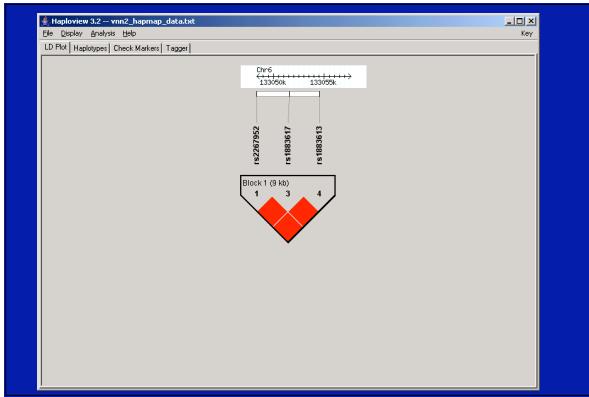
This screenshot shows a configuration dialog titled "Configure... SNP genotype data". It has sections for "Population" (set to "CEU") and "Output format" (with options "Text", "Save to Disk", and "Open directly in HaploView (NB doesn't work on all OS platforms or browser)"). Buttons include "Cancel", "Configure", and "Go". The main area contains a large block of text representing SNP genotype data in a tab-delimited format.

The screenshot shows the homepage of the Haplovew website at <http://www.broad.mit.edu/mpg/haplovew/>. It features a header with "Haplovew" and a "Tutorial" link. Below the header is a "How to use Haplovew" section with a "Get Started" button. The main content area includes sections for "How to use Haplovew", "FAQ", "Contact & License Info", and "Mailing List". A "Feedback" section at the bottom encourages users to report bugs.

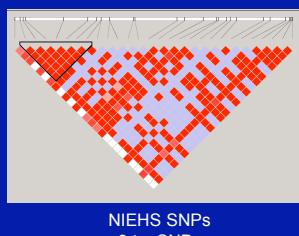
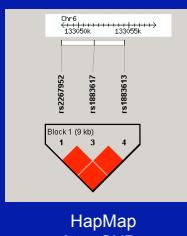
This screenshot shows the "Import HapMap Data into Haplovew" interface. It consists of two overlapping windows. The top window is titled "Welcome to Haplovew" and has three buttons: "Load genotypes (bit-age format)", "Load phased haplotypes", and "Load HapMap data". A red arrow points to the "Load HapMap data" button. The bottom window is also titled "Welcome to Haplovew" and shows file selection fields: "Genotype file: PGK\_education.hgv", "Browse", "Load phased haplotypes", and "OK" and "Cancel" buttons.

This screenshot shows the "LD Plot" interface of Haplovew 3.2. It displays a table of LD pairs between SNPs. A red arrow points to the "LD Plot" tab in the top navigation bar. The table includes columns for ID, Name, Position, Distan, Pvalue, Fdr, Haplotype, 3Dmbo, FamSize, Mendel, MAF, and Rating. At the bottom, there are controls for "Haplotype p-value cutoff" (0.0010), "Min genotype %" (75), "Max 3 mendl errors" (1), "Select All", "Minimum minor allele freq" (0.0010), and "Recompute Markers".

This screenshot shows the same LD Plot interface after data has been loaded. The table now lists 10 LD pairs. A red arrow points to the "LD Plot" tab again. The bottom controls remain the same, but the "Block 1" section now displays "TAT .708", "CGC .200", and "CAT .052".



## Variation data, LD, and tagSNPs for *VNN2* in European-Americans



## Where to Find Tagging Software

HaploBlockFinder	<a href="http://cgl.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi">http://cgl.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi</a>
Haplovew	<a href="http://www.broad.mit.edu/mpg/haplovew/">http://www.broad.mit.edu/mpg/haplovew/</a>
LDSelect	<a href="http://droog.gs.washington.edu/lSelect.html">http://droog.gs.washington.edu/lSelect.html</a>
SNPtagger	<a href="http://www.well.ox.ac.uk/~xiayi/haplotype/index.html">http://www.well.ox.ac.uk/~xiayi/haplotype/index.html</a>
TagIT	<a href="http://popgen.biol.ucl.ac.uk/software.html">http://popgen.biol.ucl.ac.uk/software.html</a>
tagSNPs	<a href="http://www-rcf.usc.edu/~stram/tagSNPs.html">http://www-rcf.usc.edu/~stram/tagSNPs.html</a>

## Haplotypes, TagSNPs, and Caveats

- Haplotypes are inferred
- Block-like structure assumed for some software
- Different block definitions
- Block boundaries sensitive to marker density
- Genotype savings may not be great (recombination)

## Common Errors in Association Studies

Bell and Cardon (2001)

- Small sample size
- Subgroup analysis and multiple testing
- Random error
- Poorly matched control group
- Failure to attempt study replication
  - e.g., Second case/control study  
Gene expression studies
  - ✓ Failure to detect LD with adjacent loci
- Over interpreting results and positive publication bias
- Unwarranted 'candidate gene' declaration after identifying association in arbitrary genetic region

## Picking SNPs Application to Association Studies Summary

- Resources available for pair-wise LD and haplotypes
- Software for tagSNP selection available
- Be aware the limitations of the approach you choose
- Replication required by several journals